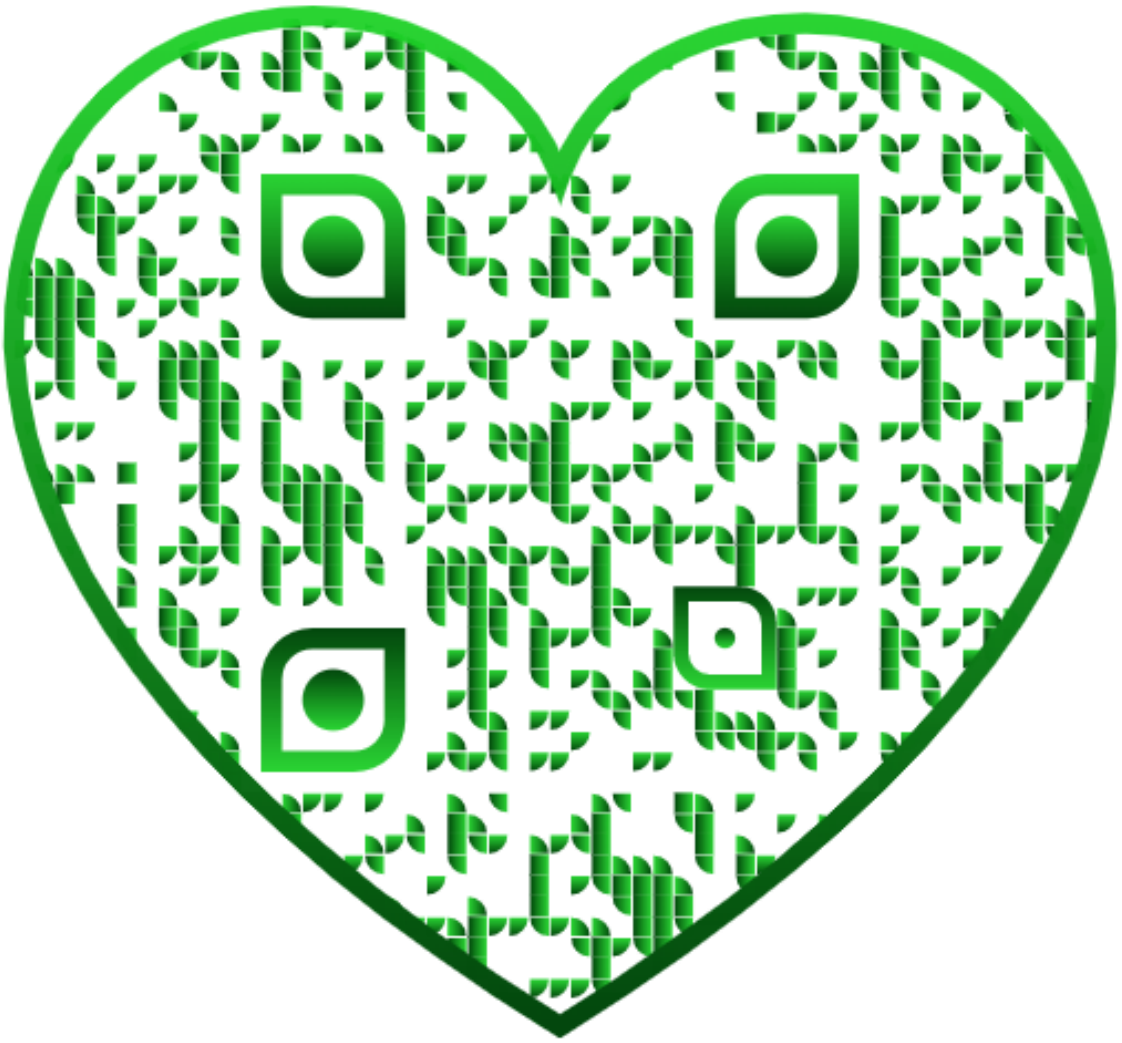
CHAPTER -7
ESTIMATION

❑Learning objectives

❑ At the end of this chapter the student will be able to:

1. Understand the concepts of sample statistics and population parameters
2. Understand the principles of estimation and differentiate between point and interval estimations
3. Compute appropriate confidence intervals for population means and proportions and interpret the findings
4. Describe methods of sample size calculation for cross – sectional studies



Inferential Statistics

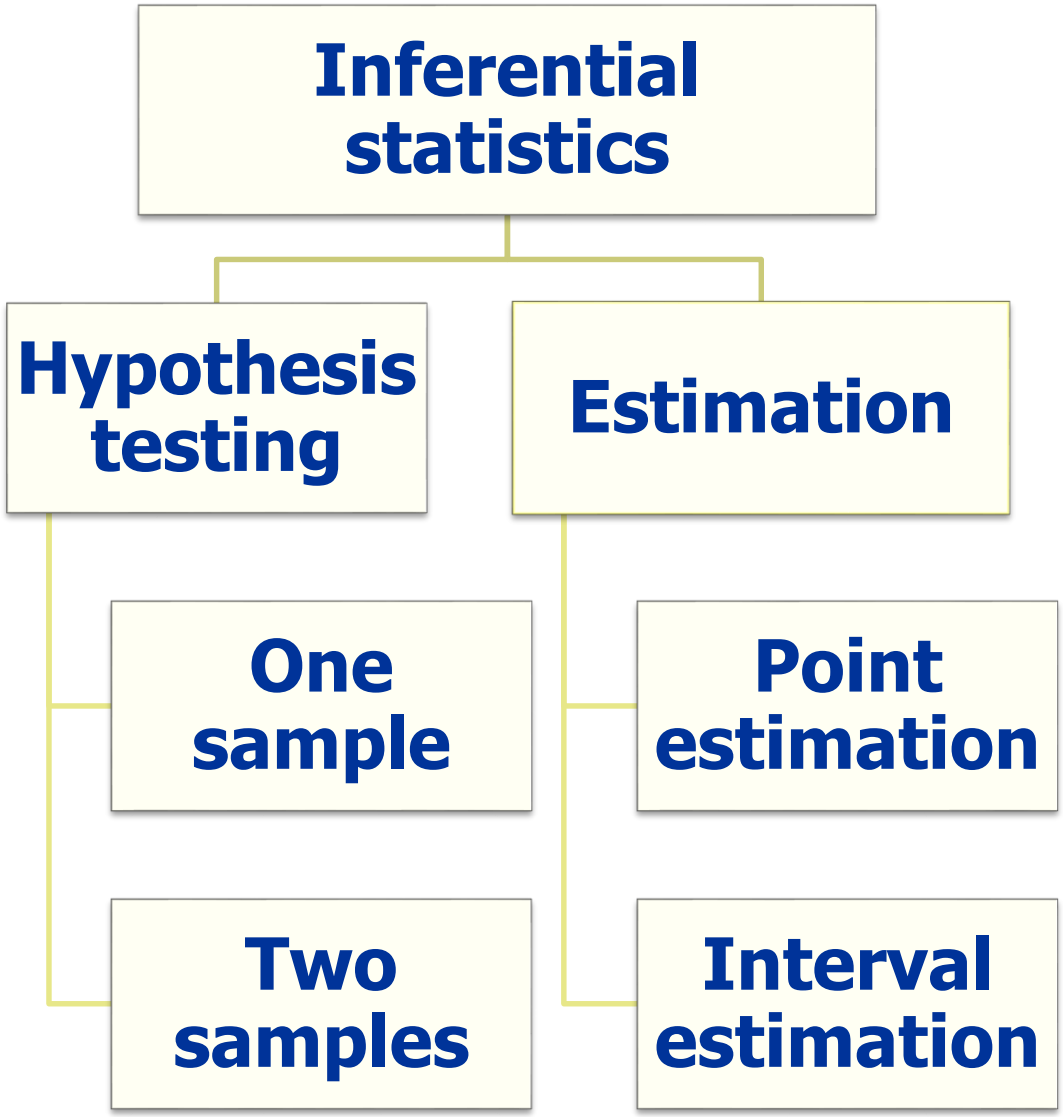
- **Inferential Statistics**: Are statistical methods used for drawing conclusions about a population based on the information obtained from the sample of observations drawn from that population

Inferential Statistics

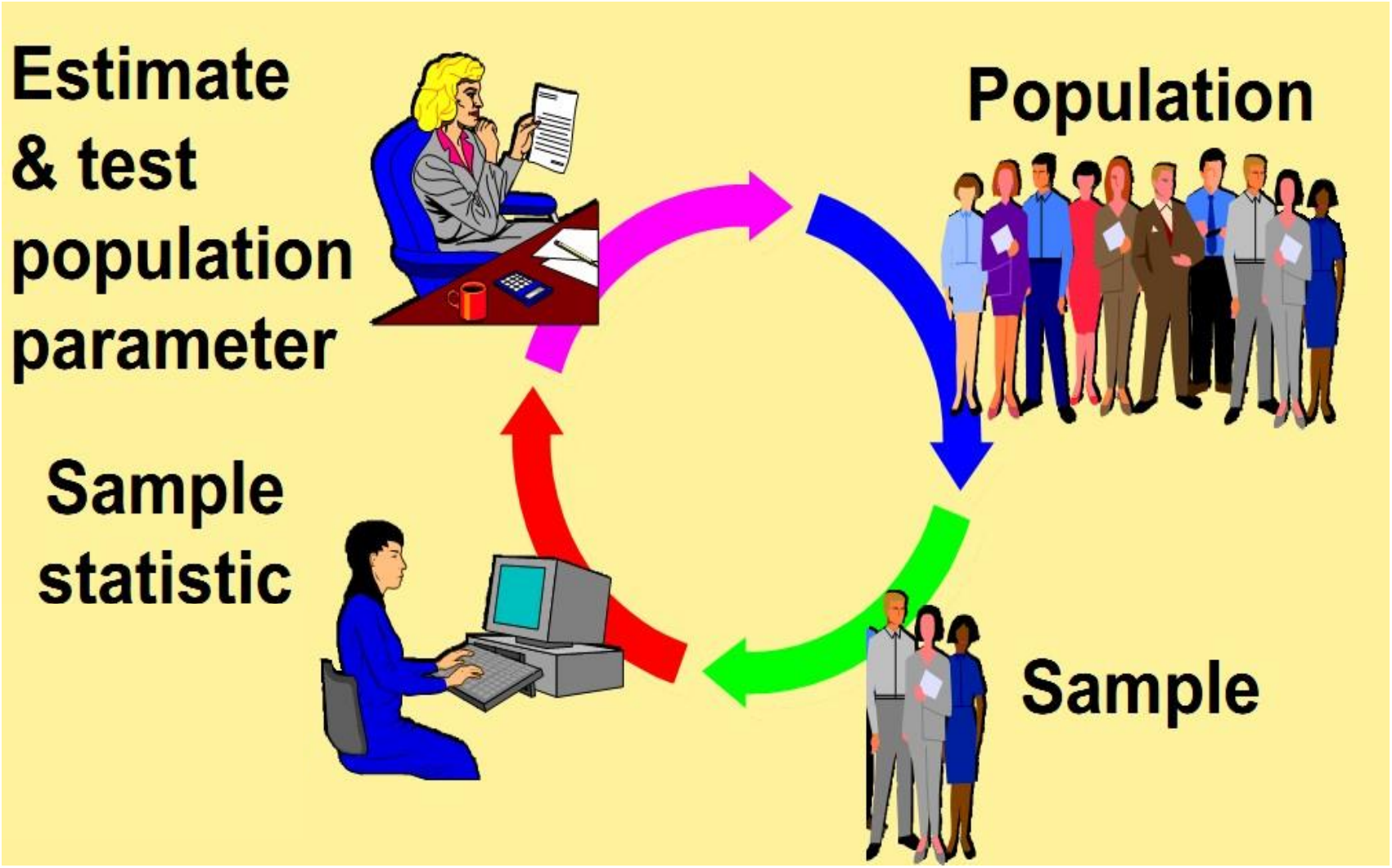
- Involves
  - Estimation
  - Hypothesis testing

- Purpose
  - Make decisions about population characteristics

Inferential Statistics



Inferential process



Statistical Estimation

- **Estimation** is the process of determining a likely value of population parameter, based on information collected from the sample.
- **Estimation** is the use of **sample statistics** to estimate the corresponding **population parameters**
- **The objective** of estimation is to determine the **approximate value** of unknown population parameter on the basis of a sample statistic

Estimation

- Definition: **A parameter** is a numerical descriptive measure of a population (μ is an example of a parameter).
- **A statistic** is a numerical descriptive measure of a sample (*X is an example of a statistic).*
- To each sample statistic there corresponds a population parameter.
- We use $\bar{X}$,S^2 , S , p, *etc. to estimate* μ, σ^2, σ, P *(or* π*), etc.*

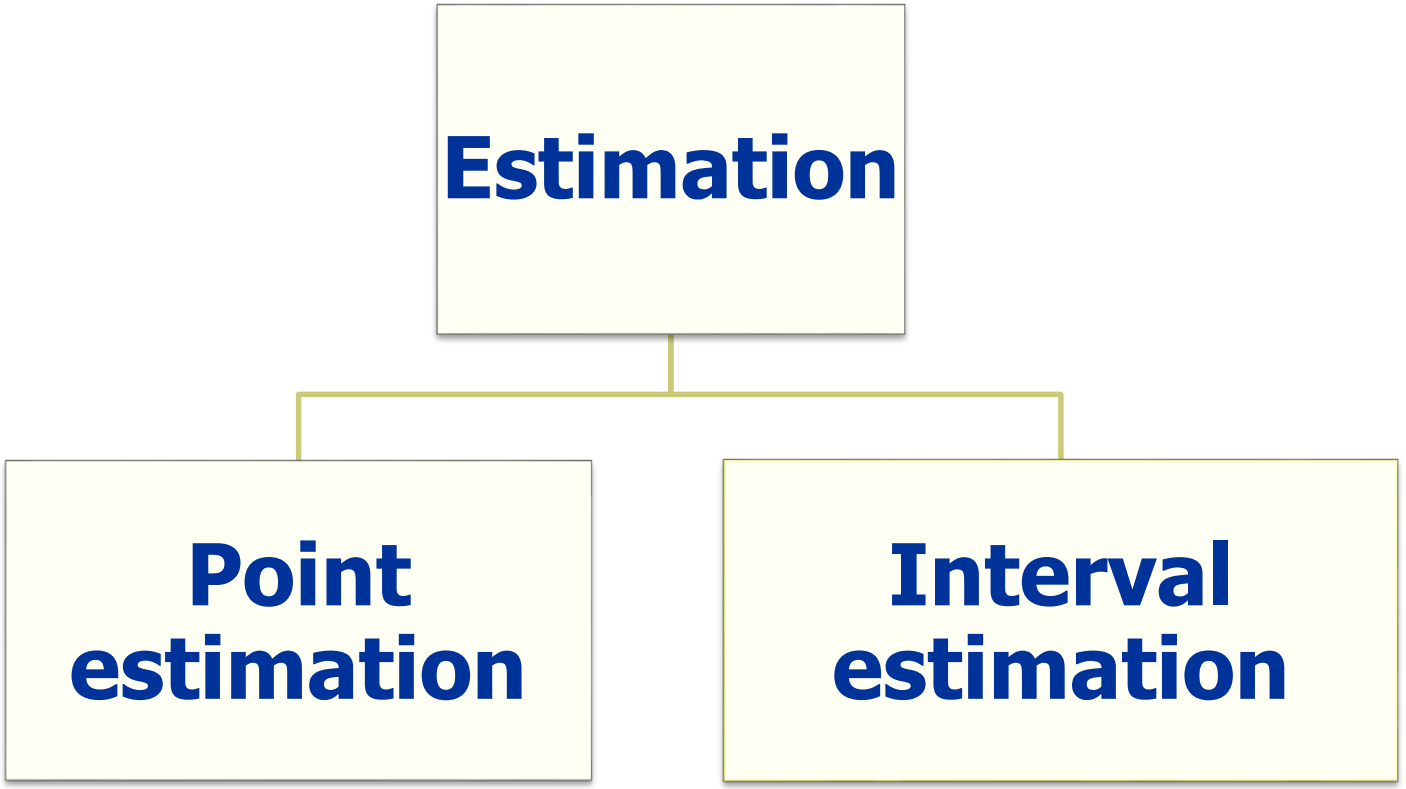
Sample statistic &their corresponding population parameter

	Statistic		Parameter
Mean:	$\bar{x}$	estimate	μ
Variance	s^2	estimate	σ^2
Standard deviation:	s	estimate	σ
Proportion:	$\hat{p}$	estimate	π

From sample

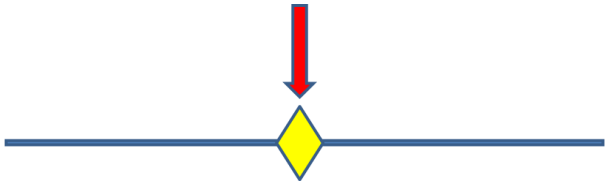
From entire population

Estimation



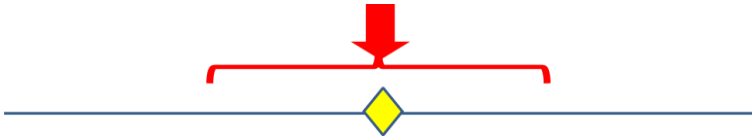
Estimation: (two types)

1. **Point estimation**



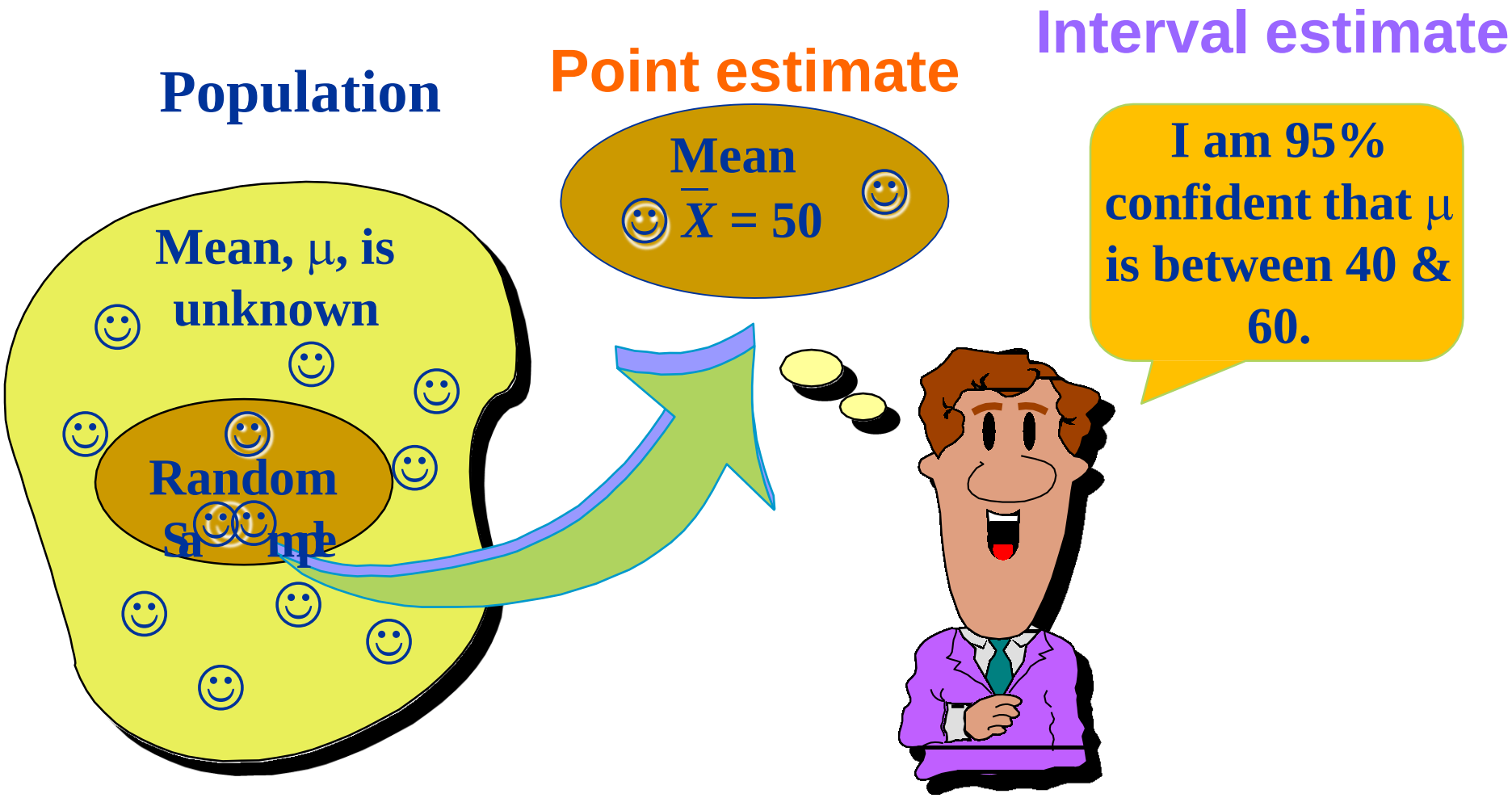
- ✓ A single numerical value used to estimate the corresponding population parameter
- ✓ Point estimation ignores **sampling error.**

2. **Interval estimation**



- ✓ Using a sample statistic to estimate a population parameter by making allowance for sample variation (error) .
- ✓ It is a statement that describes a population parameter has a value lying in between two specified limits with a certain confidence interval.

Estimation Process



Point estimation

- A single numerical value used to estimate the corresponding population parameter
- Gives little information about how close the value is to the unknown population parameter
- Example: Sample mean $\bar{X} = 3$ is **point estimate** of unknown population mean

Interval estimation

- A single-valued estimate conveys little information about the actual value of the population parameter, is not reasonable
- Assume that a sample statistic value is not exactly equal to the corresponding population parameter
- An interval estimate which locates the population parameter within an interval, with a level of confidence is needed

Marginal Error

- A value calculated from the sample is the best guess when estimating corresponding population value
- Estimate is still uncertain due to sampling error
- Marginal Error is a measure of uncertainty
- Using marginal error you can state:

Confidence interval = Estimate ± Marginal error

Confidence interval for a single mean

- Using a formula

$$\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad \Rightarrow \quad \left(\bar{x} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

Confidence Level	Critical Z
90%	1.645
95%	1.960
99%	2.576

Example (1)

- The mean reading speed of a random sample of 81 adults is 325 words per minute. Find a 90% C.I. For the mean reading speed of all adults (μ) if it is known that the standard deviation for all adults is 45 words per minute.
- **Answer** = (316.8, 333.2)
- **Interpretation;** we are 90% confident that the truth value in population is between the above CI.
- α is to be chosen by the **researcher**, most is common values of α are 0.05, 0.01, 0.001 and 0.1.

Example (2)

- ✓ A physical therapist wished to estimate, with 99% confidence, the mean maximal strength of a particular muscle in a certain group of individuals. He assume that strength scores are approximately normally distributed with a variance of 144. A sample of 15 subjects who participated in the experiment yielded a mean of 84.3, find 99% C.I _____?

Answer (76.3, 92.3)

Interpretation; we are 99% sure/confident that the mean maximal strength of a particular muscle in population is between CI (76.3, 92.3)

C.I. For The Difference Between two Population Means

- **Known variance (2 independent samples)**
- **A 100(1- α)% C.I. for μ1 - μ2 is**

$$\left(\overline{X}_1 - \overline{X}_2\right) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Example (1)

- If a random sample of 50 non-smokers have a mean life of 76 years with a standard deviation of 8 years, and a random sample of 65 smokers live mean of 68 years with a standard deviation of 9 years,

A. Find a 95% C.I. for the difference of mean lifetime of non-smokers and smokers.

B. Interpretation**?**

Interpretation; we are 95% sure/confident that the difference of mean lifetime of non-smokers and smokers in population is between CI ((4.88, 11.12 years)

Answer: A) (4.88, 11.12 years)

Example (2)

- An anthropologist who wanted to study the heights of adult men and women took a random sample of 128 adult men and 100 adult women and found the following summary results [n1= 128; n2= 100; mean1= 170cm; mean2 = 164cm; SDm=8; SDw=6]. Find point e, interval estimate and interpret the result.

- **Answer;** **Point estimate = 6cm & 95% C.I = (3.76, 8.24)**

C.I. for a Single Population Proportion

- For large sample size
- A 100(1 - α)% C.I. for π is-

$$P \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{P(1-P)}{n}}$$

- Where, p = proportion of success in a population,

1-P = proportion of failure and n = sample size

Example (1)

- An epidemiologist is worried about the ever increasing trend of malaria in a certain locality and wants to estimate the proportion of persons infected in the peak malaria transmission period. If he takes a random sample of 150 persons in that locality during the peak transmission period and finds that 60 of them are positive for malaria, then find 99% confidence intervals for the proportion of the whole infected people in that locality during the peak malaria transmission period . **Answer (0.297, 0.503)**

Example (2)

- A study on dental health practice. Of 300 adults interviewed, 123 said that they regularly had a dental checkup twice a year. What is the 95% C.I. for π(P)?

- **Answer (0.36, 0.46)**
- **Interpretation = ???**

C.I. for the difference between two population proportions

- For large sample size; (n >= 30)
- A 100(1- α)% C.I. for $\pi_1 - \pi_2$ or $(P_1 - P_2)$ is:-

$$(P_1 - P_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}$$

Example (1)

- Two hundred patients suffering from a certain disease were randomly divided into two equal groups. Of the first group, who received the standard treatment, 78 recovered within three days. Out of the other 100, who were treated by a new method, 90 recovered within three days. The physician wished to estimate the true difference in the proportions who would recovered within three days. Assume he wants to estimate 95% C.I
- **Answer (-0.22, -0.02)**
- **What does the negative sign indicates?**

Exercise :

1) The mean weight of 100 children who are 5 years old in a certain locality is found to be 14 kg. A clinician wants to know the mean weight of all the children in that locality with 95 % confidence interval, if it is known that the SD for all children is 4kg. find 95% C.I?
2) A survey conducted on a reprehensive sample of 900 newborn babies in A/A and it is found that their average weight at birth is 3.5 kg with SD of 0.5Kg estimate the wt of newborn babies in A/A at the 95% level of confidence.
3) A random sample of 100 people shows that 25 are left-handed. Find a 95% CI for the true proportion of left-handers

- Calculate CI and interpret each finding?

Sample Size Determination

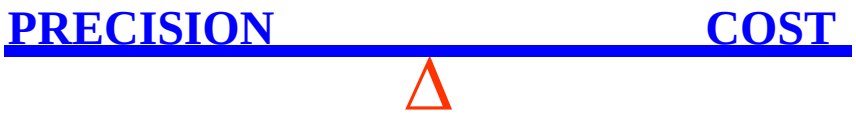
- **An essential part of planning any study is to decide how many people need to be studied and how to choose them.**

Sample Size (=n)

- ***Sample Size*: The number of study subjects selected to represent a given**
  - Study population.
  - Important to make inferences based on the findings from the sample.
  - Should be sufficient to represent the characteristics of interest of the study population.

- **Common questions:**
  - ***"How many subjects should I study?"***
  - **Too small sample = Waste of time and resources**
    = Results have no practical use
  - **Too large sample = Waste of resources**
    = Data quality compromised
    = Any small difference can be statistically significant

When deciding on sample size:



↑ **Sample size =** ↑ **Precision =** ↑ **Cost**

- **Precision is related to confidence level & CI**

Sample size determination depends on the:

- **Objective of the study**
- **Design of the study**
  - **Descriptive/Comparative**
- **Degree of precision or accuracy – the allowed deviation from the true population parameter (can be within 1% or 5%, etc)**
- **Plan for statistical analysis**
- **Degree of confidence level required, usually specified as 95%:**
  - **For example: the confidence that the proportion in**
    - **the whole population is between (p-d) and (p+d)).**

Sample size for single sample includes:

A. Sample size for estimating a single population mean
B. Sample size to estimate a single population proportion

A. Sample size for estimating a single population mean

A 100(1-α)% C.I. for μ is:

We use a sample mean as a point estimate of μ

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

α is to be chosen by the researcher, the most common values of α are 0.05, 0.01 and 0.1.

A. Sample size for estimating a single population mean

- **AIM: Estimate μ**
- **WANT: Estimate ($\bar{X}$) ± d units**

where d = Margin of error =

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

= Absolute precision

= Half of the width (w) of CI

Steps:

1. **Specify d (or w = 2d)**
2. **Use known σ^2 or estimate using s^2**

3.

$$d = z_{\alpha/2}\sqrt{\sigma^2/n}$$

$$d^2 = (z_{\alpha/2})^2 \cdot \sigma^2/n$$

estimator of the Standard error of the parameter of interest

$$n = \frac{(z_{\alpha/2})^2 \cdot \sigma^2}{d^2}$$

$$n = \frac{z^2_{\alpha/2}\ \sigma^2}{\quad}$$

Where d = e in some text books

Example(1):

1. Find the minimum sample size (n) needed to estimate the drop in heart rate (μ) for a new study using a higher dose of propranolol than the standard one. We require that the two-sided 95% CI for μ be no wider than 5 beats per minute and the sample sd for change in heart rate equals 10 beats

per minute.

$$n = (1.96)^2 10^2/(2.5)^2 = 62 \text{ patients}$$

To change the confidence level, the multiplier $(1.96)^2$ as follows: 90% CL=$(1.64)^2$, 99% CL=$(2.58)^2$

Example(2):

2. Suppose that for a certain group of cancer patients, we are interested in estimating the mean age at diagnosis. We would like a 95% CI of 5 years wide. If the population SD is 12 years, how large should our sample be?

$$n = \frac{(z_{\alpha/2})^2 \cdot \sigma^2}{d^2} = \frac{(1.96)^2(144)}{(2.5)^2} = 88.5 \approx 89$$

- **Suppose d=1**
- **Then the sample size increases**

$$n = \frac{z^2 \cdot \sigma^2}{d^2} = \frac{(1.96)^2(144)}{(1)^2} = 553.2 \approx 554$$

Exercise:

1. **A hospital administrator wishes to estimate the mean weight of babies born in the hospital. How large a sample of birth records should be taken if she/he wants a 95% CI of 0.5 wide? Assume that a reasonable estimate of σ is 2.**
   Ans: 246 birth records.

B. Sample size to estimate a single population proportion

- **Aim:** Estimate p
- **Want:** Estimate ± d units where d = Z•SE

  (95% CI of width=2d)

Steps:

1. Specify d (or w = 2d)
2. Use estimated p (use p=0.5 if no information)

3. Solve for n

$$d = z_{\alpha/2}\sqrt{\frac{pq}{n}}$$

$$d^2 = (z_{\alpha/2})^2 \cdot \frac{pq}{n}$$

$$n = \frac{(z_{\alpha/2})^2 \cdot pq}{d^2}$$

Example(1)

1. **Suppose that you are interested to know the proportion of infants who breastfed >18 months of age in a rural area. Suppose that in a similar area, the proportion (p) of breastfed infants was found to be 0.20. What sample size is required to estimate the true proportion within ±3% points with 95% confidence. Let p=0.20, d=0.03, α=5%**

$$n = \frac{(z_{\alpha/2})^2 \cdot pq}{d^2} = \frac{(1.96)^2(0.2)(0.8)}{(0.03)^2} = 683$$

- **Suppose there is no prior information about the proportion (p) who breastfeed**
- **Assume p=q=0.5 (most conservative)**
- **Then the required sample size increases**

$$n = \frac{(z_{\alpha/2})^2 \cdot pq}{d^2} = \frac{(1.96)^2(0.5)(0.5)}{(0.03)^2} = 1{,}068$$

Example(2)

2. Suppose that We wish to estimate the prevalence of asthma in an adult population with the width of the 95% confidence interval of 0.10, an accuracy of ±0.05. An estimate of the prevalence of asthma is 0.10.

Ans: 138

If we choose to half the accuracy to give a 95% confidence interval of 0.05 width: $n = 1.96^2 \times 0.1\,(1-0.1)/0.025^2 = 553$

Example(3)

3) p = 0.26 , d = 0.03 , Z = 1.96 (i.e., for a 95% C.I.)

- **Thus, the study should include at least 822 subjects.**

Points for Consideration

1. Sample size estimates might need to be adjusted to compensate for non-response rate, patient dropout or loss to follow-up, lack of compliance, etc.
2. If sampling is from a finite population of size N, then:

$$n = \frac{n_0}{\left(1 + \frac{n_0}{N}\right)}$$

where n_0 is the sample from an infinite population. When N is large in comparison to n, (i.e., $n/N \leq 0.05$), the finite population correction may be ignored.

3. Design effect for complex cluster sampling. Common values:

multiply n by 2, 3, …5.

If the above sample is to be taken from a relatively small population (sayN = 3000), the required minimum sample will be obtained from the above estimate by making some adjustments (if the population is less than 10,000 then a smaller sample size may be required.).

$$n_{final} = \frac{821.25}{1 + \frac{821.25}{3000}} = 644.7 \approx 645$$

Using A Census For Small Populations

- Use the entire population as the sample.
- Attractive for small populations (200 or less).
- Eliminates sampling error and provides data on all the individuals in the population.
- Financial considerations make this impossible for large populations.

Exercise

1) A survey of school children to determine the population of immunized children against polio, an investigator determined the maximum discrepancy b/n sample and population proportion of immunized to be 0.04, at level of confidence of 99%.further the investigator had a previous knowledge on the prevalence among children in a similar community to be 90% and the total population of school children is 800.

2) A health officer wishes to estimate the mean serum cholesterol in a population of men. From previous similar studies a standard deviation of 40 mg/100ml was reported. If he is willing to tolerate a marginal error of up to 5 mg/100ml in his estimate, how many subjects should be included in his study? (α =5%, two sided).

3) A hospital administrator wishes to know what proportion of discharged patients are unhappy with the care received during hospitalization . If 95% Confidence interval is desired to estimate the proportion within 5%, how large a sample should be drawn ?

4) a population of cancer patients has survival standard deviation of 43.4 months. If one wants to conduct a study on these populations how large sample size is needed, so that 95% of the sample mean of this size will be within ±6 months of the population mean. Population size is 480 patients.

Reading assignment

- CI for population mean: Large-sample size and when σ is unknown.
- CI for population mean: small sample size (n<30) and when σ is unknown
- **Sample size for estimating a double population mean**
- **Sample size for estimating a double population proportion**?

Next class: Hypothesis testing!!

Thanks a lot for your attention!!

Any question please !!!